HRTFs for binaural synthesis

Møller, Henrik; Hammershøi, Dorte

Published in:
Proceedings of NORSIG'98 : 3rd IEEE Nordic Signal Processing Symposium, June 8-11, 1998, Vigsø, Hanstholm, Denmark

Publication date:
1998

HRTFs FOR BINAURAL SYNTHESIS

Henrik Møller, Dorte Hammershøi

Acoustics Laboratory, Aalborg University
Fredrik Bajers Vej 7 B4, DK-9220 Aalborg Ø, Denmark

ABSTRACT

This paper presents the method of binaural synthesis for use in creation of three-dimensional sound images in virtual reality and multimedia applications. Special attention is given to the selection of head-related transfer functions (HRTFs), and how this choice affects the further processing.

A brief introduction to spatial hearing is given, followed by a division of the sound transmission to the eardrum. The selection of HRTFs for binaural synthesis is discussed. It is then described how the binaural synthesis is carried out using information on sound transmission in the room as input. Examples of headphone characteristics are given, and it is described how the correct eardrum signals are obtained by headphone reproduction.

1. INTRODUCTION

The hearing has two inputs: sound pressure at the two eardrums. From these inputs the hearing creates an image of the acoustical surroundings, determines direction and distance to sound sources, extracts sound from a single source in noisy surroundings etc.

The fact that the input to the auditory system consists of the two eardrum sound pressures only, is utilized in the **binaural recording** technique. Sound is recorded in the ears of a human subject (or an acoustical mannequin, an artificial head) and played back through headphones.

The three dimensional effect is overwhelming. The acoustics of the recording room is precisely reproduced, sources can be perceived in all directions, including up and down, and they can come close to the listener, down to a few centimeters from the ear.

Thus, if a listener is given the correct sound at the two eardrums, he may be given an auditory perception that differs from the perception corresponding to his actual physical situation.

In binaural recording, the role of the head (whether human or artificial) is to transform each sound wave into two sound pressures, one for each ear. If sufficient knowledge is available about the transmission to the ears for sound from "all" directions (as many as we can discriminate), then it is possible to program a computer to simulate the transmission. The art of artificially creating the eardrum signals is called **binaural synthesis**.

Because of the possibility of rendering audible situations which do not exist in real life, binaural synthesis has a great potential in virtual reality and multimedia applications.

A brief summary of spatial hearing and the sound transmission from a free field to the eardrum is given in Sections 2 and 3 of this paper. Section 4 mentions the sound transmission in the room, information on which is needed as a prerequisite for the binaural synthesis. Calculation of the room transmission is not an issue in the present paper. The binaural synthesis process is described in Section 5.

The head-related transfer functions, the HRTFs, for the synthesis may be measured on human, individually or non-individually, or they may originate in measurements on artificial heads. Possible sources for HRTFs are discussed in Section 6.

The computer generated signals are typically reproduced by headphones, because these readily provide the necessary channel separation. To obtain the correct eardrum signals the synthesis must be supplemented by a correction filter, an equalization. Section 7 deals with headphone performance on human listeners and the needed equalisation.

The material presented in this paper was obtained in investigations which are reported more thoroughly in [1]-[10]. Some practical aspects of binaural synthesis, e.g. HRTF filter representation, spatial resolution, update rate and latency are given in a parallel paper [11].

2. SOUND TRANSMISSION TO THE EARDRUM

The sound transmission to each of the eardrums depends on direction to the sound source. Various effects like reflection, diffraction, shadowing, dispersion, interference and resonance are involved in a complex acoustical system formed by the body, head, pinna, ear canal and eardrum. The hearing localizes using interaural differences in level and time as well as coloration of the sound signal.

Figure 1 shows the sound transmission to the eardrums of one subject for sound coming from the left side. Left frame shows the transmission given in the time domain as impulse responses, and the right frame shows

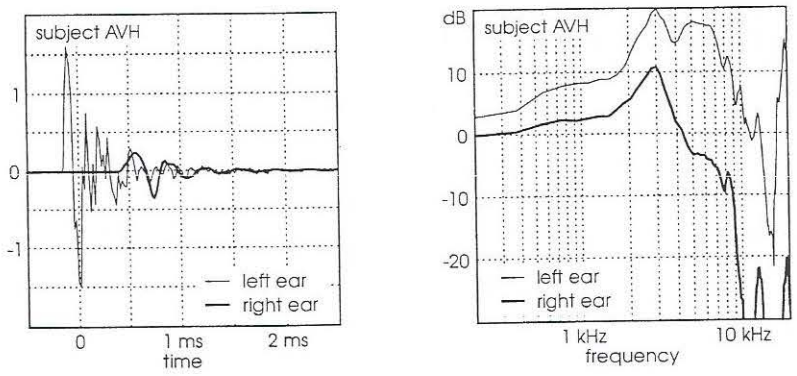

Figure 1. Sound transmission to the eardrum given in the time domain (left frame) and the frequency domain (right frame). One subject, sound incidence from left side.

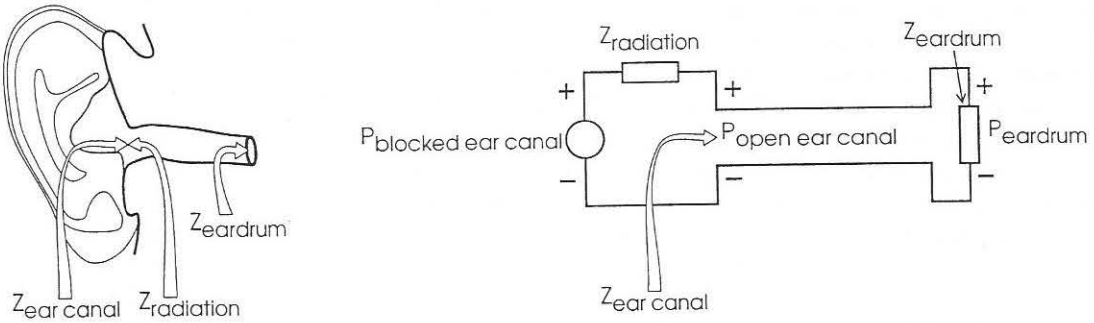

Figure 2. Sound transmission throught external ear. Sketch of anatomy (left) and analog model (right).

the transmission given in the frequency domain as amplitude responses.

The sound reaches the left ear approximately 0.6 ms before it reaches the right ear. The sound level in the right ear is lower than in the left ear, and especially the high frequencies are attenuated.

If the sound arrives from a source in the median plane, the sound transmission is almost the same to the two ears. Only a coloration - nearly identical in the two ears - serves as a cue in localization, and it may be difficult to hear the exact direction.

3. SPLITTING UP THE TRANSMISSION

Despite the fact that it is the objective in binaural synthesis to control the eardrum signals in the final reproduction, it has proven useful and more convenient to let the synthesis simulate the transmission to a point more distal in the ear canal. This is most easily understood by using a model of the transmission of sound to the eardrum. The model is given by the diagram in Figure 2.

The transmission outside the ear canal is represented by a Thevenin equivalent circuit, consisting of the impedance seen from the ear canal into the free air $Z_{radiation}$ and the generator $P_{blocked\ ear\ canal}$. "Blocked" refers to the "open circuit" situation, which is obtained by blocking the ear canal, thereby rendering the volume velocity zero.

If $P_{reference}$ denotes sound pressure at the center position of the head, but with the subject absent, then the sound transmission can be divided into three parts in the following way:

$$\frac{P_{eardrum}}{P_{reference}} = \frac{P_{blocked\ ear\ canal}}{P_{reference}} \cdot \frac{P_{open\ ear\ canal}}{P_{blocked\ ear\ canal}} \cdot \frac{P_{eardrum}}{P_{open\ ear\ canal}} \quad (1)$$

The first term is a **head-related transfer function** (HRTF). The second term is the **pressure division**

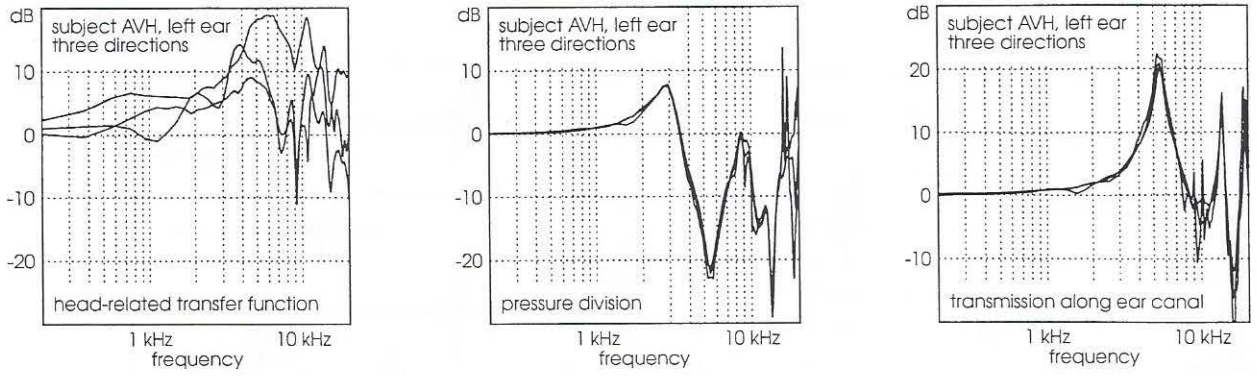

Figure 3. Parts of sound transmission to the eardrum shown for three directions: head-related transfer function (left), pressure division (center), transmission along ear canal (right).

$\frac{Z_{ear\ canal}}{Z_{ear\ canal} + Z_{radiation}}$, while the last term describes the **transmission along the ear canal**.

The three terms are shown for one subject and three angles of sound incidence in Figure 3. The way the sound reaches the ear canal does not affect the transmission within the ear canal. As a consequence, only the first term, the HRTF, depends on direction of sound incidence. The two remaining terms are the same, regardless of direction.

Thus, the HRTFs describe the directional dependent transmission of sound from the free field to the blocked ear canal. In the binaural synthesis described in Section 5, HRTFs are used to compute the blocked ear canal sound pressures. How this synthesis can secure the correct sound pressure at the eardrums by proper equalization of the headphone is described in Section 7.

Due to anatomical differences, all elements of the sound transmission to the eardrum are highly individual. The term head-related transfer function is sometimes defined in such a manner that it includes the pressure division and possibly also the transmission along the ear canal (or a part of it). However, full spatial information is included at the blocked ear canal, and the smallest effect of interindividual variation is seen here, since the inclusion of any additional transmission will add variation.

4. ROOM TRANSMISSION

Before the binaural signals can be computed, the physical transmission in the room from source to listener must be known. For each sound source, the resulting sound at the listener's position can be described as a number of attenuated and delayed sound waves reaching the listener from various directions: One wave reaching the listener directly, some first order reflections, i.e. sound waves which have been reflected in one room surface, some second order reflections, and a large number of higher order reflections (in theory an infinite number, but for practical purposes it is possible to use a finite number).

In the present presentation, the *calculation* of the room transmission is not considered part of the binaural synthesis, and it is not covered by the present paper. Of course, the *result* of the calculation constitutes an important input to the synthesis.

5. BINAURAL SYNTHESIS

In the binaural synthesis, pressure in the two ears are computed, given descriptions of all incoming sound waves at the position of the listener. For this purpose HRTFs are used. The HRTFs are most often used in the time domain, that is as head-related impulse responses (HRIRs).

The blocked ear canal pressures for the two ears $p_{left}(t)$ and $p_{right}(t)$ resulting from a single sound wave $s(t)$ can be obtained by convolution with the two sides $HRIR_{left}(t)$ and $HRIR_{right}(t)$ of the HRIR for the appropriate direction:

$$\begin{aligned} p_{left}(t) &= HRIR_{left}(t) * s(t) \\ p_{right}(t) &= HRIR_{right}(t) * s(t) \end{aligned} \qquad (2)$$

If the sound field consists of N sound waves, each described by the direction $\angle(i)$ and the time signal $s_i(t)$, then the resulting sound pressures at the blocked ear canals can be found by summation:

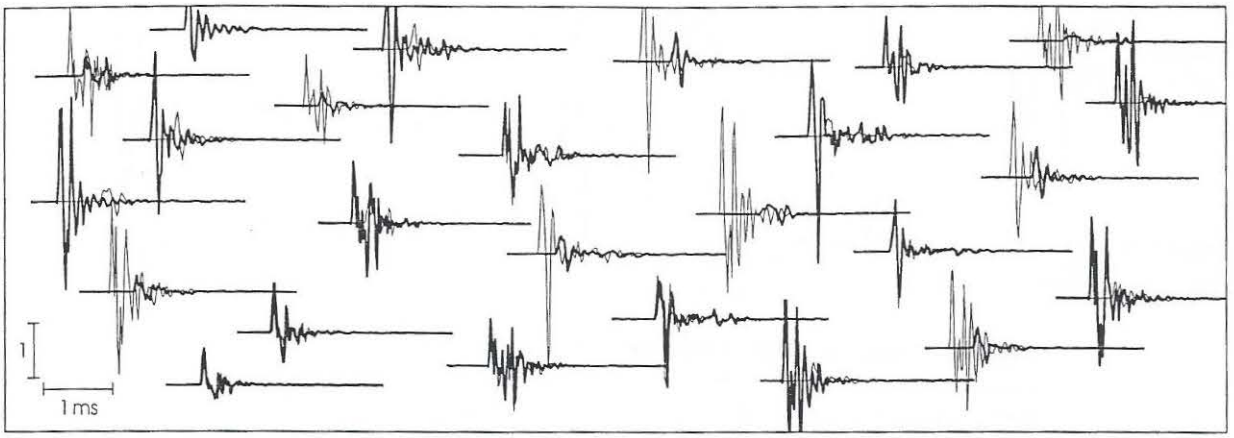

Figure 4. Examples of head-related impulse responses, HRIRs. Left and right part in heavy and thin lines, respectively.

$$p_{left}(t) = \sum_{i=1}^{N} HRIR_{left,\angle(i)}(t) * s_i(t)$$
$$p_{right}(t) = \sum_{i=1}^{N} HRIR_{right,\angle(i)}(t) * s_i(t) \quad (3)$$

$HRIR_{left,\angle(i)}$ and $HRIR_{right,\angle(i)}$ are the two sides of the HRIR for direction $\angle(i)$.

For binaural synthesis, the computer should hold a database of HRIRs. Each HRIR has a duration of 1-2 ms. For a 48 kHz sampling frequency this corresponds to 48-96 taps in an FIR filter. Figure 4 shows examples of HRIRs with the left part given as heavy lines and the right part as thin lines.

Head-related transfer functions can be split into a minimum phase part, a linear phase part (that is a delay), and an all-pass phase part. An example is given in Figure 5.

In binaural synthesis it is essential to include the minimum phase part and the linear phase part. At present it is unknown, what effect the all-pass phase part has on the sound quality.

If the various sound waves described by their time signal $s_i(t)$ originate in the same source signal $s(t)$, then

$$s_i(t) = RIR_i(t) * s(t) \quad (4)$$

where the $RIR_i(t)$ describes the room transmission for the i'th transmission path. $RIR_i(t)$ will consist of delay and attenuation corresponding to the propagated distance in combination with a filtering from the walls or other items hit on the way. A possible directional characteristic of the source will also be included. The sum $\sum_{i=1}^{N} RIR_i(t)$ is known as the *room impulse response* (RIR).

Equation (3) can now be rewritten:

$$p_{left}(t) = \sum_{i=1}^{N} HRIR_{left,\angle(i)}(t) * RIR_i(t) * s(t)$$
$$p_{right}(t) = \sum_{i=1}^{N} HRIR_{right,\angle(i)}(t) * RIR_i(t) * s(t) \quad (5)$$

Those parts of the summations in equation (5) which depend on the transmission path constitute each their part of the *binaural room impulse response* (BRIR).

$$BRIR_{left}(t) = \sum_{i=1}^{N} HRIR_{left,\angle(i)}(t) * RIR_i(t)$$
$$BRIR_{right}(t) = \sum_{i=1}^{N} HRIR_{right,\angle(i)}(t) * RIR_i(t) \quad (6)$$

Equation (5) can now be written

$$p_{left}(t) = BRIR_{left}(t) * s(t)$$
$$p_{right}(t) = BRIR_{right}(t) * s(t) \quad (7)$$

The part of the BRIR which is significant for the auditory impression has a duration in the same order of magnitude as the reverberation time for the particular room. The number of taps in a 48 kHz implementation of the FIR filters of equation (7) may thus amount to 30-100,000, or - in the case of large rooms or concert halls - even more.

At present, commercial hardware is available that is able to perform the convolution in real time. On the other hand, it is assumed that only the direct sound and the early reflections must be calculated accurately, and that other methods, for instance statistical methods, may be used to calculate late reflections and the last part of the BRIR, the reverberation. It is also believed that late parts

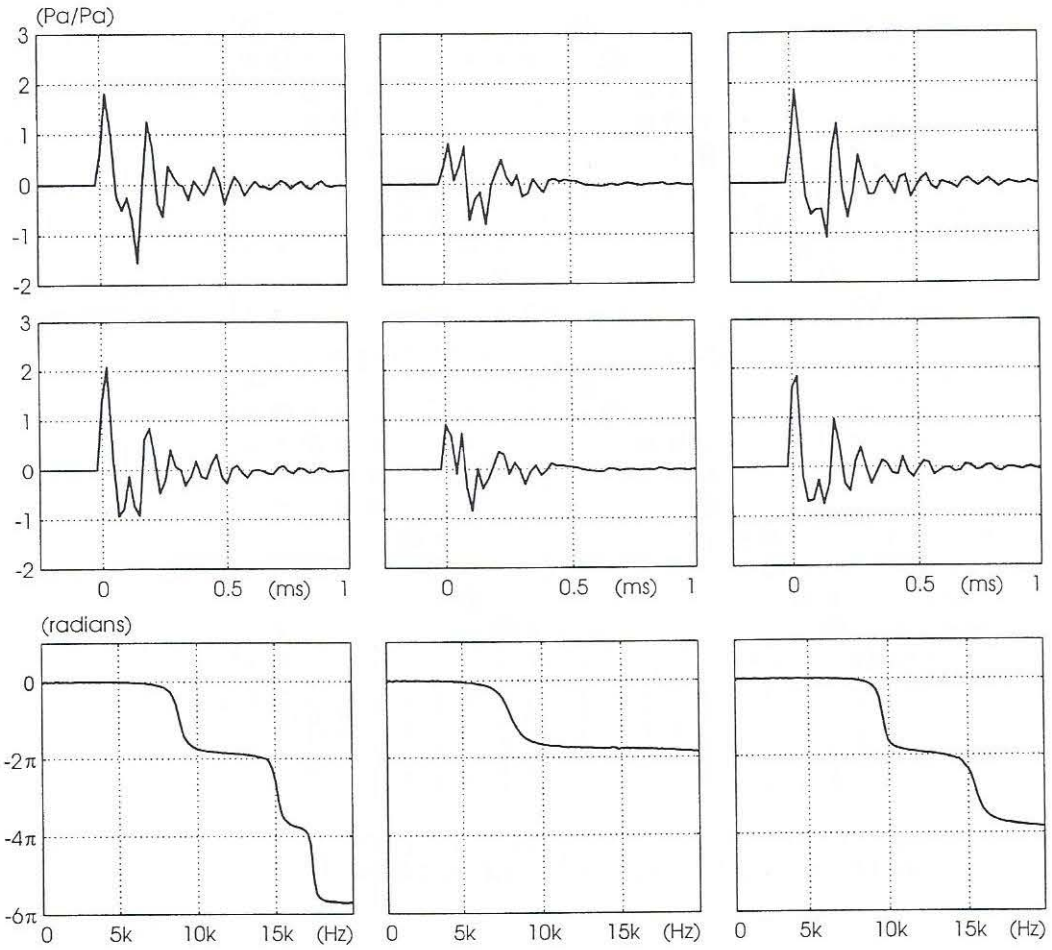

Figure 5. Three examples (each their column) of HRIRs disregarding linear phase (upper row), their minimum-phase counterparts (middle row), and their all-pass phase (lower row).

of the BRIR may be common to a specific room, thus only demanding update of the early parts in dynamic systems. A clarification of these issues is aimed at in laboratories worldwide, since simplifications are considered a precondition for the commercial succes of binaural synthesis for low-cost multimedia applications.

6. SELECTION OF HRTFs

The geometry of humans varies considerably, and ideally the HRTFs used in the binaural synthesis (and the headphone transfer function used for the equalization) should originate from the particular listener. This is, however, not possible in practical applications. Most conveniently, it should be possible to use the same binaural synthesis for all listeners.

A very logical solution would be to measure HRTFs on an artificial head, since artificial heads are constructed with the objective of simulating the acoustics of an average human.

In order to - among other things - explore the possibility of using the same binaural signal for all people, a number of experiments were carried out at our laboratory. The localization performance of 20 subjects was studied when they listened:

- in real life,
- to binaural recordings made in their own ears,
- to binaural recordings made in the ears of other humans selected randomly,
- to binaural recordings made in the ears of a carefully selected typical human (same for all listeners), and
- to binaural recordings made with artificial heads.

Note that recordings were used rather than synthesized binaural signals. This leaves out possible errors from the room modelling.

The results for median plane sound sources are given in Figure 6. In each frame stimulus direction is

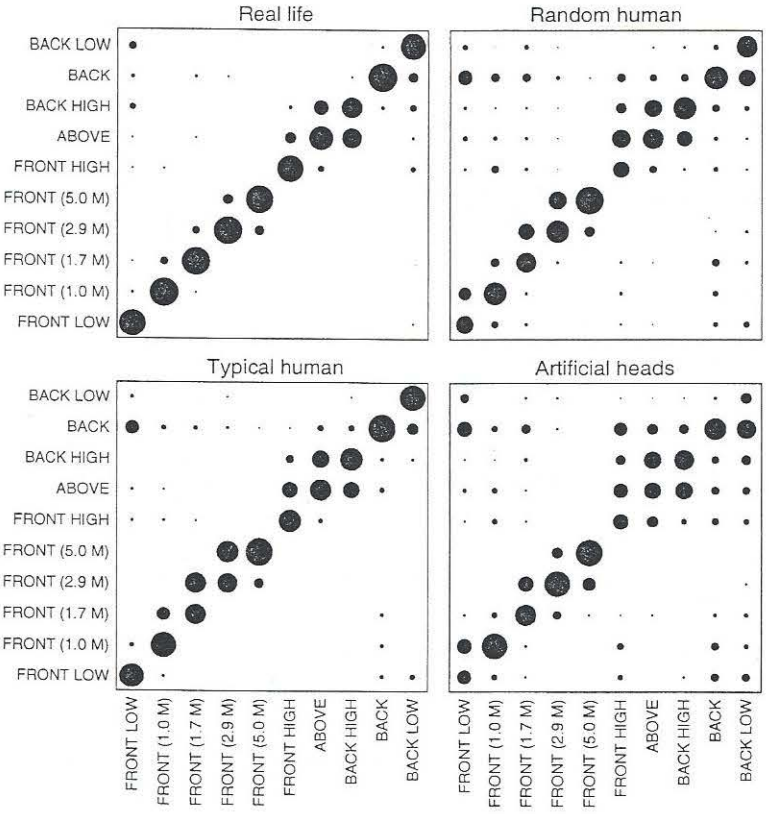

Figure 6. Median-plane performance of 20 subjects in real life and with three types of binaural recordings.

given at the horizontal axis and response direction at the vertical axis. The area of each circle indicates number of responses for the particular combination of stimulus and response. Correct responses are at the diagonal. (Results from listening to recordings made in the subject's own ears are nearly identical to those of real life listening and therefore not shown).

The localization performance is quite good in real life, although not perfect (upper left frame). Much more errors are seen, when listening to binaural recordings made in the ears of other humans, chosen at random (upper right frame). It proved possible to find a typical human, the recordings from whom resulted in much better performance, in fact a performance not much inferior to that of real life (lower left frame). The largest number of localization errors were seen with recordings from artificial heads (lower right frame).

In this figure, results from all artificial heads are pooled, and differences between heads are concealed. Figure 7 shows the percentage of median-plane errors for each make of head put into the ranking of human heads. Despite the intention with the artificial heads, they are all in the bad half of the humans.

A possible recommendation on this background would be to use HRTFs measured in the ears of a carefully selected human subject. HRTFs from artificial heads cannot be recommended at present.

7. REPRODUCTION WITH HEADPHONES

The binaural synthesis described in the preceding sections simulates the real life sound transmission only to the blocked ear canal and not the entire way to the eardrum. Assuming that the remaining transmission to the eardrum is the same during headphone listening and in real life, then the headphone should have a flat frequency response measured at the blocked ear canal.

Figure 8 shows frequency responses of three headphones, each measured at the blocked ear canal of 40 subjects. The responses are far from being flat, and equalization is needed. Individual variations are clearly seen, and individual equalization may be relevant.

The above assumption about the remaining transmission being the same during headphone listening and in real life can be verified by splitting up the transmission into the pressure division and the transmission along the ear canal.

The transmission along the ear canal in the headphone situation is physically the same and therefore identical to that of the real life situation. Only the

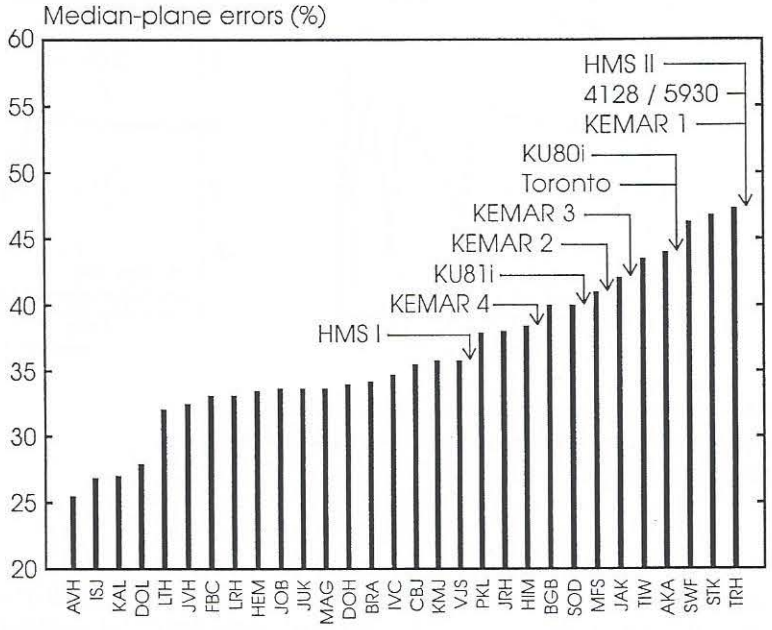

Figure 7. Median-plane performance with recordings from human heads ("recording head" given by initials). Performance with artificial heads indicated by arrows.

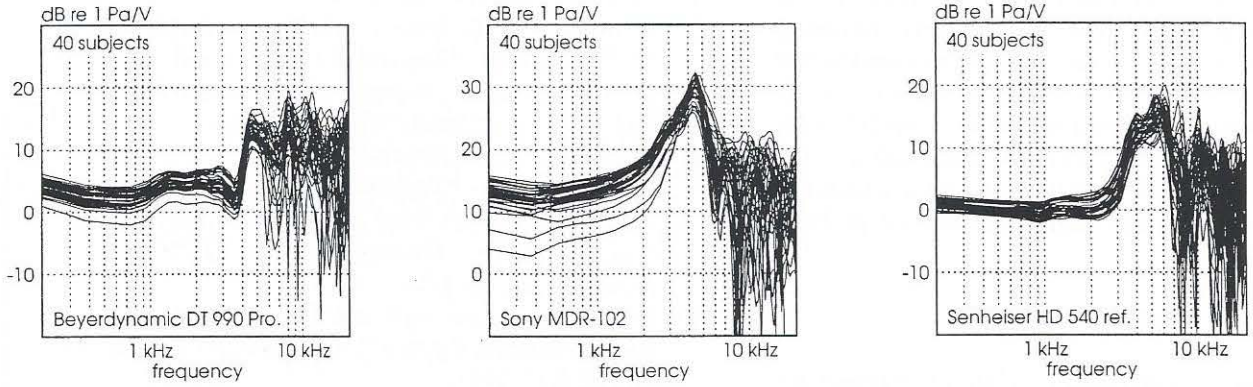

Figure 8. Frequency response of three sample headphones measured at the blocked ear canal of 40 human subjects.

pressure division is changed to $\frac{Z_{ear\ canal}}{Z_{ear\ canal} + Z_{headphone}}$. It differs from that of the real life listening situation, since the radiation impedance is replaced by $Z_{headphone}$, the acoustical impedance of the headphone as seen from the ear canal. If $Z_{headphone}$ and $Z_{radiation}$ are identical, or if they are both small compared to $Z_{ear\ canal}$, then the pressure division will be the same in the two situations.

Figure 9 shows pressure divisions measured on one subject in real life and when listening to three commercial headphones. Only minor differences are seen between real life and with headphones.

8. DISCUSSION

It was mentioned in Section 3 that the sound pressure at the blocked ear canal contains full spatial information, and it was further argued that it had the smallest effect of interindividual variation. The consequence of this is that HRTFs determined at the blocked ear canal have a more general applicability, with respect to the use for listeners other than the subject, which the HRTFs were determined for.

Apart from that, the blocked ear canal measurements also offer a number of practical advantages. The most immediate is that the microphone

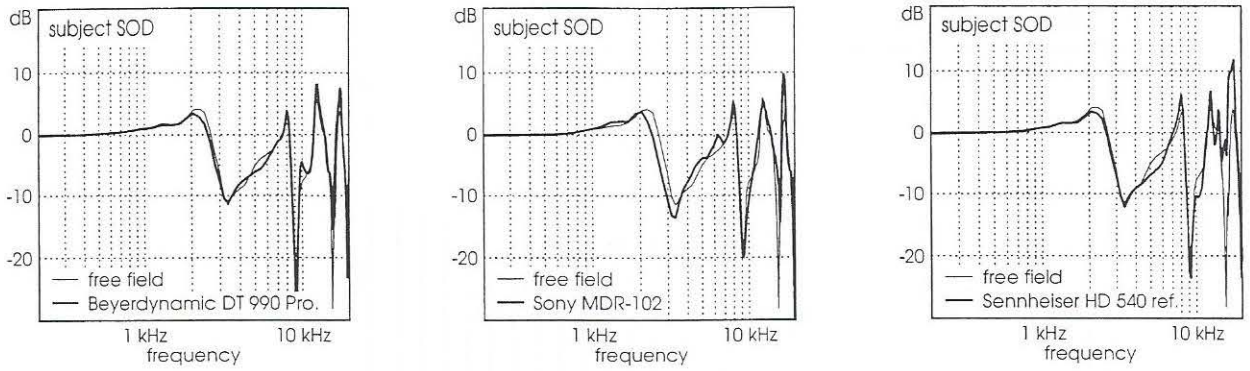

Figure 9. Pressure division in free field (thin line) and with three sample headphones (heavy line). Measurements on one human subject.

mounting becomes relatively easy, especially compared to the critical and often unpleasant measurements at the eardrum. Since the microphone can be mounted in the blockage, it can be larger and provide a better sensitivity than the probe microphones typically used for eardrum measurements.

One advantage which should also be stressed is that using blocked ear canal HRTFs, no measurements need ever be made at the eardrum. This has been overlooked by some, who believe that the transmission from blocked entrance to eardrum must be determined once for all. This is not correct. As stated in Section 7, the filters for correcting the headphone characteristics shall also be determined by measurements at the blocked entrance, and the measurements at the eardrum are thus completely avoided.

9. REFERENCES

[1] H. Møller, "Fundamentals of binaural technology", *Appl. Acoust.*, Vol. 36, No. 3/4, pp. 171-218, 1992.

[2] D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal" *J. Acoust. Soc. Am.*, Vol. 100, No. 1, pp. 408-427, July 1996.

[3] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-related transfer functions of human subjects", *J. Audio Eng. Soc.*, Vol. 43, No. 5, pp. 300-321, May 1995.

[4] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Transfer characteristics of headphones measured on human ears", *J. Audio Eng. Soc.*, Vol. 43, No. 4, pp. 218-232, April 1995.

[5] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings ?", *J. Audio Eng. Soc.*, Vol. 44, No. 6, pp. 451-469, June 1996.

[6] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen, "Using a Typical Human Subject for Binaural Recording", in preparation for the *J. Audio Eng. Soc.*, preliminarily reported in [7].

[7] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen, "Using a Typical Human Subject for Binaural Recording", *Proc. 100th Audio Eng. Soc. Conv.*, Copenhagen, Denmark, May 11-14 1996, preprint 4157, pp. 1-18.

[8] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen, "Evaluation of artificial head recording systems", in preparation for the *J. Audio Eng. Soc.*, preliminarily reported in [9].

[9] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen, "Evaluation of artificial head recording systems", *Proc. 100th Audio Eng. Soc. Conv.*, Munich, Germany, March 22-27 1997, preprint 4404, pp. 1-32.

[10] H. Møller, "Interfacing Room Simulation Programs and Auralisation Systems", *Appl. Acoust.*, Vol. 38, pp. 333-347, 1993.

[11] D. Hammershøi and J. Sandvad, "Using binaural synthesis for auditory virtual environments", in *Proc. IEEE Nordic Signal Processing Symposium, NORSIG '98*, June 6-8 1998, Vigsø, Denmark, pp. 189-192.